Provedor de Serviços para Análise de Interação

Samuel Alves Basílio
UFJF
Rua José Kelmer, S/n – Juiz de Fora - MG
Brasil – CEP 36036-330
alvesbasilio@gmail.com

Eduardo Barrére
UFJF
Rua José Kelmer, S/n – Juiz de Fora - MG
Brasil – CEP 36036-330
eduardo.barrere@ice.ufjf.br

RESUMO
Este artigo apresenta a arquitetura de um Provedor de Serviços para Análise de Interação dos usuários da Tv Digital Interativa (TVDI). O Provedor é parte das especificações do GingaAiyê, um projeto que aborda aplicações não convencionais para TVDI, seguindo os padrões do Middleware Ginga.

Categories and Subject Descriptors
H.5.4 Hypertext/Hypermedia

General Terms
Algorithms, Management, Standardization.

Keywords
Ginga, Tv Digital, Interação.

1. INTRODUÇÃO
Em 2007 a transmissão da Televisão Digital aberta no Brasil foi iniciada, tendo como um grande objetivo governamental a inclusão social[1]. Além dos ganhos sociais, os ganhos tecnológicos (imagem, vídeo e interatividade) e compõem um ambiente repleto de oportunidades para desenvolvedores, publicitários e jornalistas, entre outros.

No caso do Brasil, o middleware escolhido para ser utilizado no ambiente de Tv Digital é o Ginga[2]. A partir da especificação do middleware, vários grupos de pesquisas vêm desenvolvendo novas funcionalidades, que podem ser incorporadas a ele para suprirem demandas que não fazem parte do padrão estabelecido.

Entre essas demandas estão as estabelecidas pelo projeto GingaAiyê, que busca incorporar funções não convencionais como controle de interatividade, publicidade e criação de documentos no lado do telespectador[3]. Para tal, foi desenvolvido um Componente de Captura de Interação do Usuário (CCIU), que registra todas as interações do telespectador para posterior análise local ou remota. Estas interações podem ser desde as clássicas, troca de canais e volume, até as interações oriundas da manipulação das aplicações e dispositivos possíveis em um ambiente de Tv digital [4].

Este trabalho apresenta um Provedor de Serviços de Interação (PSAI), responsável por receber os dados capturados pelo módulo CCIU e realizar o tratamento e disponibilização desses dados para potenciais clientes. O artigo apresenta inicialmente o ambiente no qual o PSAI está inserido. Posteriormente são apresentadas e detalhadas as suas funcionalidades. É apresentada uma aplicação de prova de conceito do PSAI e o artigo é finalizado com as considerações finais e trabalhos futuros.

2. AMBIENTE DE TV DIGITAL
Para que a recepção do sinal digital aconteça é necessário que o usuário tenha um aparelho Set-Top Box (STB). Este aparelho permite que os aplicativos desenvolvidos para Tv Digital sejam executados na casa do telespectador, conforme as especificações do Ginga.

No ambiente de televisão digital interativa brasileiro, as emissoras enviam o sinal digital, via broadcast, para os telespectadores que, por sua vez assistem aos programas e interagem com as aplicações através de uma STB. Quando a aplicação necessita enviar dados para as emissoras e demais servidores, é necessário que o usuário tenha um canal de retorno, por exemplo um link de internet. Nesse ambiente é que se insere o PSAI, vide Figura 1.

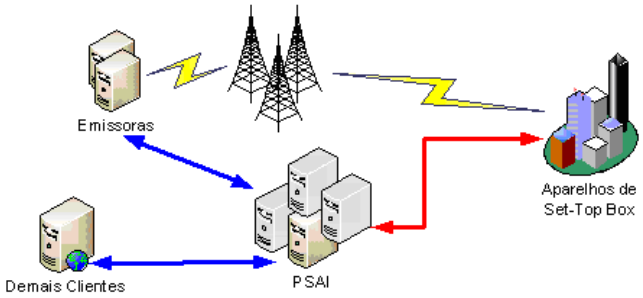

Figura 1 - Ambiente de TVDI com um PSAI

Conforme apresentado na Figura 1, o PSAI exerce o papel de intermediário entre as informações capturas em cada STB e as emissoras de televisão ou demais empresas interessadas nas interações dos usuários, como empresas de medição de audiência, fabricantes de equipamentos e empresas de publicidade, entre outras. Como exemplo de uso do PSAI, uma emissora pode saber se uma determinada aplicação, desenvolvida para ser executada em conjunto com um programa, está ou não sendo acessada pelos telespectadores.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
Conference'10, Month 1–2, 2010, City, State, Country.
Copyright 2010 ACM 1-58113-000-0/00/0010…$10.00.

3. PSAI

A ideia central é que o PSAI funcione como um receptor das interações ocorridas e armazenadas nas diversas STBs, armazene essas informações e possa prover serviços que disponibilizem esses dados aos diversos interessados. Neste contexto, são funcionalidades, desenvolvidas como módulos, do PSAI:

- Receber e armazenar as interações dos usuários;
- Realizar o processo de ETL (*Extract, Transform, and Load*) [5] para a geração de Data Warehouses (DW);
- Fornecer acesso aos dados dos DW para os clientes;
- Fornecer aplicações web.

Para realizar tais funcionalidades, o PSAI pode ser desde um simples servidor ou um cluster de servidores capaz de suportar uma grande quantidade de STBs e clientes, seguindo os critérios clássicos de escalabilidade de hardware para servidores web.

3.1 Recepção de Dados

O módulo CCIU, desenvolvido no Projeto GingaAyê, armazena na STB toda interação que o telespectador realiza, via controle remoto ou outro dispositivo de interação. Posteriormente essas interações são enviadas a um PSAI, previamente configurado na STB, através de um Serviço Web. A Figura 2 ilustra este processo de transferência de dados entre uma STB e um PSAI.

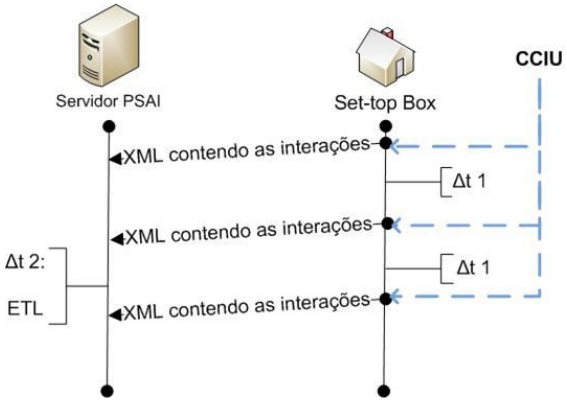

Figura 2 – Comunicação PSAI x STB

Conforme configurado no STB do telespectador, a cada intervalo de tempo (Δt1), por exemplo a cada 24 horas, um arquivo XML é enviado ao PSAI contendo todas as interações realizadas desde o último envio. Um exemplo simplificado de arquivo XML enviado pode ser visto na Figura 3.

```
<?xml version="1.0" encoding="UTF-8" ?>
<watchTV country="Brazil" startDate="2010-06-18T08:30:00.0" endDate="2010-06-
   18T17:25:00.0">
<head>
  <location zip="36035330" lat="-21.455100" long="-43.210100" />
  <user birth="01-02-1972" genre="male"> ……. </user>
  <resources> …..- </resources>
</head>
<interaction type="channelChange" time="2010-06-18T08:30:00.0">
  <key code="CH_UP" />
  <channel code="05" name="Globo">
    <program code="160" name="Copa do Mundo 2010 na Globo- Jogo 1"
    category="Esportes" age="10" />
  </channel>
 </interaction>
 <interaction> ….. </interaction>
</watchTV>
```

Figura 3 – Modelo Dimensional Simplificado do DW

O documento é divido em duas partes principais, o cabeçalho (<head>) que armazena informações sobre o usuário e a STB e as interações (<interaction>), que armazena todos os tipos de interações realizadas na STB, pelos diversos dispositivos de interação na casa do telespectador (interações explícitas) ou por ações internas da Tv, como finalização de um programa e início do próximo (interações implícitas).

Esses arquivos são enviados para um serviço web, fornecido pelo PSAI, que por sua vez armazena todos os dados do XML em um banco de dados relacional[6]. O motivo de armazenamento de todos os dados do XML, sem nenhum processamento inicial, é o fato de poder haver inúmeras STBs enviando dados para o PSAI, o que torna importante a rapidez de execução do serviço web, ou seja, a ausência de processamento dos dados durante a recepção.

3.2 Processo de ETL

A cada período de tempo Δt2 (ilustrado na Figura 2), um processo de ETL é executado a fim fazer a carga do DW com as últimas interações recebidas pelo PSAI. Este período de tempo pode variar conforme alguns fatores, principalmente:

- Volume de dados recebidos;
- Capacidade de processamento do servidor/cluster;
- Quantidade de DW diferentes e consequentemente quantidade de processos de ETL a serem executados; e
- Necessidades temporais dos clientes em consultar os dados;

3.3 Fornecimento de Acesso aos Dados

Os clientes de um PSAI podem ter acesso aos dados conforme diversas regras de negócio estabelecidas, mas pensando em termos de tecnologia, os dados podem ser fornecidos através de:

- Serviço web que permite realizar consultas específicas;
- Exportação de dados via HTTP ou FTP;
- Aplicações web que permitem aos clientes terem acesso às pesquisas sobre um determinado DW.

Esta é a parte mais flexível de um PSAI, pois pode ser modelada conforme o tipo de negócio ao qual o mesmo for projetado.

4. Implementação

O PSAI foi implementado utilizando Java, banco de Dados MySQL e serviços web baseados em TomCat e Axis. Toda a parte de recebimento dos arquivos XML e armazenamento dos mesmos em um banco de dados relacional está implementada e funcional. Como exemplo de aplicação a ser fornecida pelo PSAI foi implementado um DW, cujo modelo dimensional tem como atividade central a interação realizada em uma STB. A Figura 4 apresenta a um esquema simplificado do modelo dimensional.

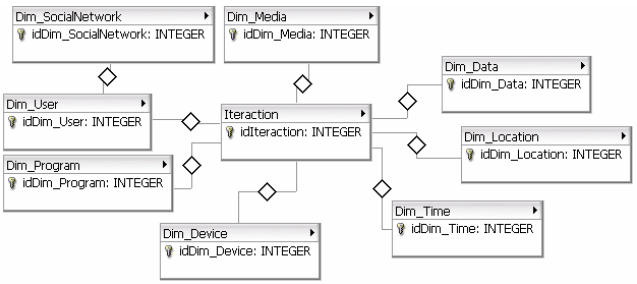

Figura 4 – Modelo Dimensional Simplificado do DW

No processo de ETL para esse DW, a maioria dos dados é oriunda da base de dados relacional do PSAI, somente as dimensões Data (Dim_Data) e Localização (Dim_Location) são preenchidas com dados obtidos de fontes externas ao PSAI. A dimensão Data deve conter informações como: feriados, finais de semanas de demais dados que trazem peculiaridade para cada dia do ano. Já a dimensão Localização deve permitir que sejam especificados: rua, bairro, região da cidade, nome da cidade, estado e país; a partir das informações contidas na *tag Location* (cabeçalho do arquivo XML enviado pelo STB).

A execução do processo de ETL é disparada a cada período de 24 horas (Δt2, ilustrado na Figura 2), mas pode ser ativada diretamente pelo sistema, sempre que o administrador achar necessário. Após a execução do ETL os dados estarão atualizados na base de dados do DW e poderão ser consultados pelos clientes do PSAI. Na atual versão de implementação as seguintes consultas estão disponíveis:

- Quantidade de aparelhos online em um determinado período e em um determinado canal/programa;
- Quantidade de aparelhos online em um determinado período e em um determinado canal/programa durante o fim de semana que se encontra num determinado período (vide Figura 4);
- Quantidade de aparelhos online em um determinado período e em um determinado canal/programa durante um feriado específico;
- Aplicações executadas em um determinado período (vide Figura 5);
- Quantidade de aparelhos online numa determinada faixa horária;
- Quantidade de aparelhos online numa determinada faixa horária em um determinado canal;
- Aplicações executadas em uma determinada faixa horária;

Outra funcionalidade é a exportação de dados, armazenados em um determinado período, do DW para um cliente, através de download de um arquivo XML ou via serviço web.

Figura 4 – Audiência da Emissora num Final de Semana

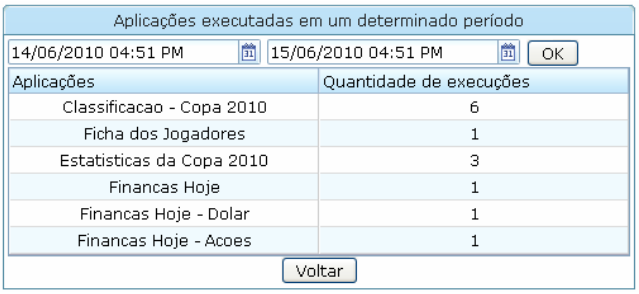

Figura 5 – Aplicações Executadas num período

Outras pesquisas estão sendo desenvolvidas para estender o protótipo, conforme demandas específicas dos projetos que fazem uso do PSAI como mecanismo de *feedback* das aplicações em execução no ambiente de TVDI.

5. CONCLUSÃO

O desenvolvimento da especificação e implementação do PSAI levou em conta aspectos essenciais de software, deixando as partes de arquitetura e escalabilidade de hardware e desempenho para a próxima etapa. No projeto atual se buscou garantir a especificação do PSAI e o desenvolvimento de um conjunto de aplicações suficientemente robustas para servir de prova de conceito. Neste aspecto, o PSAI se encontra em fase final de validação, onde dados reais (XML enviados pelas STBs) serão inseridos para testar a eficiência do provedor e para que possíveis erros possam ser encontrados e corrigidos.

6. AGRADECIMENTO

Agradecemos ao CTIC pelo apoio e financiamento desta pesquisa e ao LApIC da UFJF por fornecer a infraestrutura para sua execução.

7. REFERÊNCIAS

[1] Brasil, 2003, Decreto-lei n. 4.901, de 26 de novembro de 2003. “Institui o Sistema Brasileiro de Televisão Digital - SBTVD, e dá outras providências”. Diário Oficial da República Federativa do Brasil. Brasília, 27 de nov. 2003. Seção 1, Pág. 7.

[2] Ginga, 2008, Ginga Digital TV Middleware Specification: Padrões Ginga. In: http://www.ginga.org.br/padroes.html.

[3] Pimentel, M. G. et al. 2008. Ubiquitous end-user live editing of interactive multimedia programs. In: Webmedia’08 - XIV Brazilian Symposium on Multimedia and the Web. p. 123–129.

[4] Costa, R., Moreno, M., Soares, L. 2009. Ginga-ncl: Suporte a múltiplos dispositivos. In: Webmedia’09 - XV Brazilian Symposium on Multimedia and the Web.

[5] Kimball, R., CASERTA, Joe. 2004. Data Warehouse Etl Toolkit. Editora John Wiley Professional, 1ª Edição.

[6] Basilio, S. C. A., Barrere, E. 2009. Mecanismo para Armazenamento das Interações dos Usuários de Set-top Box em uma Ferramenta para Análise de Interação In: Simpósio Internacional de Televisão Digital, SIMTVD.